Draft version May 14, 2020
Typeset using LaTeX **preprint** style in AASTeX63

A Data-Driven Technique Using Millisecond Transients to Measure the Milky Way Halo

E. Platts,[1] J. Xavier Prochaska,[2,3] and Casey J. Law[4]

[1] *High Energy Physics, Cosmology & Astrophysics Theory (HEPCAT) group, Department of Mathematics and Applied Mathematics, University of Cape Town, South Africa*

[2] *Department of Astronomy & Astrophysics, UC Santa Cruz, USA*

[3] *Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU; WPI), The University of Tokyo, Japan*

[4] *Department of Astronomy and Owens Valley Radio Observatory, California Institute of Technology, Pasadena, CA 91125, USA*

(Dated: May 2020)

ABSTRACT

We introduce a new technique to constrain the line-of-sight integrated electron density of our Galactic halo $\mathrm{DM_{MW,halo}}$ through analysis of the observed dispersion measure distributions of pulsars $\mathrm{DM_{pulsar}}$ and fast radio bursts $\mathrm{DM_{FRB}}$. We model these distributions, correcting for the Galactic interstellar medium, with kernel density estimation—well-suited to the small data regime—to find lower/upper bounds to the corrected $\mathrm{DM_{pulsar}}/\mathrm{DM_{FRB}}$ distributions: $\max[\mathrm{DM_{pulsar}}] \approx 7 \pm 2\,(\mathrm{stat}) \pm 9\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$ and $\min[\mathrm{DM_{FRB}}] \approx 63^{+27}_{-21}\,(\mathrm{stat}) \pm 9\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$. Using bootstrap resampling to estimate uncertainties, we set conservative limits on the Galactic halo dispersion measure $-2 < \mathrm{DM_{MW,halo}} < 123\,\mathrm{pc\,cm^{-3}}$ (95% c.l.). The upper limit is especially conservative because it may include a non-negligible contribution from the FRB host galaxies and a non-zero contribution from the cosmic web. It strongly disfavors models where the Galaxy has retained the majority of its baryons with a density profile tracking the presumed dark matter density profile. Last, we perform Monte Carlo simulations of larger FRB samples to validate our technique and assess the sensitivity of ongoing and future surveys. We recover bounds of several tens $\mathrm{pc\,cm^{-3}}$ which may be sufficient to test whether the Galaxy has retained a majority of its baryonic mass. We estimate that a sample of several thousand FRBs will significantly tighten constraints on $\mathrm{DM_{MW,halo}}$ and offer a valuable complement to other analyses.

1. INTRODUCTION

In the early universe the majority of baryons resided in a cool, diffuse plasma, which is predicted to have collapsed into sheetlike and filamentary structures that make up the intergalactic medium (IGM). Around the time of structure formation, dark matter collapses into halos, pulling baryons with it. As the gas falls inwards, it is shock-heated to form a hot, diffuse plasma, known as halo gas or the circumgalactic medium (CGM). Approximately 10% of the gas cools and falls into the center of the halo to form stars and the interstellar medium (ISM; e.g. White & Rees 1978).

Comparing the baryonic mass fraction detected for galaxies ($M_b/M_{\rm halo}$) to the cosmic mean (Ω_b/Ω_m), however, reveals a baryonic deficit (e.g. Dai et al. 2010). The missing baryons may have been ejected back into the IGM before forming stars or perhaps have yet to be detected (e.g. Prochaska et al. 2011; Booth et al. 2012). In the latter scenario, the CGM presents itself as a possible refuge.

This issue holds for the CGM of our Galaxy. While it is evident that its stars and ISM correspond to $\lesssim 25\%$ of the baryonic mass available to a halo with mass $M_{\rm halo} = 10^{12.2}\mathrm{M}_\odot$ (the current estimate; Boylan-Kolchin et al. 2013), the mass and distribution of gas within our Galactic halo are not well determined even despite our close proximity. The key observables that constrain the Galactic CGM include soft X-ray emission from the plasma (Henley et al. 2010), X-ray and UV absorption-lines of oxygen ions (Faerman et al. 2017; Kovács et al. 2019), density constraints from ram-pressure stripping of the Large Magellanic Cloud (LMC; Salem et al. 2015), and dispersion measure (DM) observations from pulsars towards the LMC (Manchester et al. 2006). These have provided valuable constraints for models of the Galactic halo, but still allow for large variations in the mass and spatial extent of the gas (Fang et al. 2013; Bregman et al. 2018; Faerman et al. 2013; Prochaska & Zheng 2019).

A primary challenge to assessing the Galactic CGM is that the gas is too diffuse (especially at large radii) to be imaged directly. Furthermore, the absorption-line measurements (e.g. O VI and O VII) require substantial ionization and/or metallicity corrections to infer the to-

tal gas. In this respect, the DM measurements towards the LMC provide the most direct probe of the ionized gas, yet it lies at only $\approx 1/4$ the virial radius r_{200} of the Galaxy. Ideally, one would prefer to record DM measurements to r_{200} and also across the sky to search for asymmetries in the halo gas distribution. Just such an opportunity is now afforded (albeit with caveats, as we will discuss) by the transients known as fast radio bursts (FRBs).

FRBs are the population of ∼millisecond chirps of bright radio emission at approximately GHz frequencies discovered serendipitously (Lorimer et al. 2007) and now pursued in earnest with dedicated projects and facilities (Caleb et al. 2016; Law et al. 2018; CHIME/FRB Collaboration et al. 2018; Kocz et al. 2019). Recorded in each FRB event is its DM value $\mathrm{DM_{FRB}}$. The majority greatly exceed estimates for our Galactic ISM and CGM, lending strong statistical support that FRBs have an extragalactic origin (Petroff et al. 2019; Cordes & Chatterjee 2019). This inference has been confirmed by a small but growing set of FRBs localized to $\approx 1''$ and then shown to reside in a distant galaxy (Tendulkar et al. 2017; Bannister et al. 2019; Ravi et al. 2019; Prochaska et al. 2019; Marcote et al. 2020). As a result, the community now recognizes FRBs as a viable tool to probe ionized gas across the universe, e.g. to conclusively detect the so-called "missing" baryons of the present-day universe (Fukugita et al. 1998; Macquart 2018).

Owing to its integral nature, $\mathrm{DM_{FRB}}$ includes contributions from all of the electrons along the sightline: the intergalactic medium, gas in distant Galactic halos, the ionized gas of the system hosting the FRB, and our Milky Way. Indeed, the host and Galaxy contributions ($\mathrm{DM_{host}}$, $\mathrm{DM_{MW}}$) are frequently considered a "nuisance" to proposed analyses of the cosmic web. In this manuscript, however, we view them as a highly desired signal, i.e. a new opportunity to constrain the Galactic CGM.

There are two primary challenges that this paper addresses: how to use pulsars and FRBs to probe the dispersion measure of Galactic halos, and how to do so with a limited data set. The first problem is addressed by constraining the DM contribution of the MW halo to the total observed DM of pulsars and FRBs.

For the second challenge, only ∼ 100 FRBs have been observed to date; this necessitates techniques that are well suited to dealing with small data sets. We propose the use of standard kernel density estimation (KDE; Silverman 1986) and asymmetric, variable-bandwidth KDE (Chen 2000; Hoffmann & Jones 2015) to find probability density functions (PDFs) of the DM distribution of pulsars and of FRBs, respectively. Other density estimation techniques are explored—namely, density estimation using field theory (DEFT; Kinney 2014, 2015; Chen et al. 2018) and a generalized extreme value (GEV), but prove to be insufficient (see § B and § C for details). From the PDFs one can estimate the maximum MW halo DM given by pulsars, and the minimum MW halo and host halo DM given by FRBs. This infers constraints on the DM of the MW CGM and part of the host CGM.

We measure a MW halo DM of $63^{+27}_{-21}\,(\mathrm{stat}) \pm 9\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$, corresponding to a 1σ confidence detection. The precision of this measurement is limited by the FRB sample size and we predict a robust detection of the MW halo with the incorporation of FRB detections anticipated in the coming year. The techniques presented here will make the best precision and least ambiguous measurement of the MW halo in several years with samples of 10^4 FRBs.

The paper is structured as follows. § 2 outlines the core concepts of this work. § 3 details the density estimation techniques used in the analysis. The methodology and results are presented in § 4, where § 4.2 provides constraints based on observed data and § 4.3 provides an analysis based on simulations. The results and implications are discussed in § 5, and conclusions are summarized in § 6.

2. THE FRAMEWORK

Pulsars and FRBs are both millisecond radio transients. The former lie in the disk of the MW galaxy and the latter are extragalactic. Since the group velocity of the electromagnetic wave depends on the free electron density (n_e) along the path of propagation, the arrival time of the transient signal is extended. This spread is described by the dispersion measure:

$$\mathrm{DM} = \int \frac{n_e ds}{1+z}. \tag{2.1}$$

DMs can therefore be used to study the distribution of baryons along the line of sight between a transient source and an observer.

Figure 1 shows a schematic of how electrons are distributed relative to pulsars and FRBs. Galactic halos are assumed to be devoid of radio transients, but contain a significant column density of electrons. Pulsars have been detected predominantly in the Galactic disk or nearby globular clusters[1] (Manchester et al. 2005). Those with known distance have been used to create detailed models of the electron density distribution of the

[1] The more distant pulsars purported to reside in the Magellanic clouds (e.g., Ridley et al. 2013) are excluded from this analysis.

Milky Way disk (Cordes & Lazio 2002, 2003; Gaensler et al. 2008; Yao et al. 2017). In the following we adopt both the NE2001[2] and YMW16[3] algorithms.

If we assume FRBs are distributed throughout their host galaxies and throughout space, then the lowest $\mathrm{DM_{FRB}}$ values set a bound on the electron column density associated with the halos of the Milky Way and the typical host galaxy. This measurement is the focus of the manuscript. Table 1 provides a summary of the notation used in this paper.

Quantity	Description
$\mathrm{DM_{pulsar}}$	The total DM measurement of a pulsar
$\mathrm{DM_{FRB}}$	The total DM measurement of an FRB
$\mathrm{DM^{\delta}_{ISM}}$	DM from a fraction of the Galactic ISM
$\mathrm{DM_{ISM}}$	Total sightline DM for the Galactic ISM
$\mathrm{DM_{MW,halo}}$	DM of all gas in our Galactic halo
$\mathrm{DM^{\delta}_{MW,halo}}$	DM from a fraction of gas in our Galactic halo
$\mathrm{DM_{IGM}}$	DM from the IGM (gas between halos)
$\mathrm{DM_{cosmic}}$	DM from all cosmic gas (IGM+halos)
$\langle\mathrm{DM_{cosmic}}\rangle$	Average DM from all cosmic gas
$\mathrm{DM_{host}}$	DM from FRB host galaxy halo

Table 1. Notation

2.1. *Constraints from Pulsars*

We consider

$$\mathrm{DM_{pulsar}} = \mathrm{DM^{\delta}_{ISM}} + \mathrm{DM^{\delta}_{MW,halo}} \quad , \tag{2.2}$$

with $\mathrm{DM^{\delta}_{ISM}}$ the ISM contribution and $\mathrm{DM^{\delta}_{MW,halo}}$ the halo contribution. We then define an ISM-corrected quantity $\Delta\mathrm{DM_{pulsar}}$, which subtracts the total ISM contribution along the pulsar sightline,

$$\Delta\mathrm{DM_{pulsar}} = \mathrm{DM_{pulsar}} - \mathrm{DM_{ISM}} \quad . \tag{2.3}$$

Most pulsars have unknown distances yet are expected to lie predominantly in the Galactic disk, with a scale height of 100 pc (Faucher-Giguère & Kaspi 2006). Therefore, $\mathrm{DM_{ISM}}$ is generally larger than $\mathrm{DM_{pulsar}}$ and the majority of $\Delta\mathrm{DM_{pulsar}}$ values will be negative. Any positive values could be attributed to the halo, and therefore the maximum $\Delta\mathrm{DM_{pulsar}}$ yields a lower limit:

$$\mathrm{DM_{MW,halo}} > \max\left[\Delta\mathrm{DM_{pulsar}}\right] \quad . \tag{2.4}$$

Such an analysis must allow for uncertainties in the modeling of $\mathrm{DM_{ISM}}$, but for high Galactic latitudes these uncertainties are expected to be less than $10\,\mathrm{pc\,cm^{-3}}$.

[2] Available in Python at https://github.com/FRBs/ne2001

[3] Available in Python at https://github.com/telegraphic/pygedm

2.2. *Constraints from FRBs*

$\mathrm{DM_{FRB}}$ has contributions from the ISM, the MW halo, cosmic gas, and the FRB host galaxy,

$$\mathrm{DM_{FRB}} = \mathrm{DM_{ISM}} + \mathrm{DM_{MW,halo}} + \mathrm{DM_{cosmic}} + \mathrm{DM_{host}} \quad . \tag{2.5}$$

Similar to the pulsars, we define an ISM-corrected measure:

$$\Delta\mathrm{DM_{FRB}} = \mathrm{DM_{FRB}} - \mathrm{DM_{ISM}} \quad . \tag{2.6}$$

From the full distribution of $\Delta\mathrm{DM_{FRB}}$, we will examine the lowest values on the expectation that these have lower $\mathrm{DM_{cosmic}}$ contributions. For reference, an FRB at $z = 0.03$ (e.g. Marcote et al. 2020) has an average $\langle\mathrm{DM_{cosmic}}\rangle \approx 25\mathrm{pc\,cm^{-3}}$.

The lowest values of $\Delta\mathrm{DM_{FRB}}$ should also reflect the lowest combinations of $\mathrm{DM_{MW,halo}}$ and $\mathrm{DM_{host}}$. We expect significant variations in the latter both due to the distribution of host galaxy masses and also from variations in the FRB location within the galaxy. We express $\mathrm{DM^{min}_{host}}$ as the minimum of this distribution which may be 10 to several tens $\mathrm{pc\,cm^{-3}}$.

Regarding variations in $\mathrm{DM_{MW,halo}}$, galaxy formation models tend to predict a nearly spherical distribution of gas, especially beyond the inner halo (but see Yamasaki & Totani (2020) which includes a non-spherical component). Spherically symmetric models of our Galaxy yield less than $10\,\mathrm{pc\,cm^{-3}}$ variations in $\mathrm{DM_{MW,halo}}$ even though the Sun is located off-center (Prochaska & Zheng 2019). In the following, we will assume a single $\mathrm{DM_{MW,halo}}$ unless otherwise discussed. One recovers

$$\mathrm{DM_{MW,halo}} + \mathrm{DM^{min}_{host}} = \min\left[\Delta\mathrm{DM_{FRB}}\right] \quad , \tag{2.7}$$

and therefore

$$\mathrm{DM_{MW,halo}} < \min\left[\Delta\mathrm{DM_{FRB}}\right] \quad . \tag{2.8}$$

3. KERNEL DENSITY ESTIMATION

KDE is a non-parametric technique that estimates an unknown density by constructing a kernel at each data point and summing their contributions. Owing to their shapes, the distributions of $\Delta\mathrm{DM_{pulsar}}$ and $\Delta\mathrm{DM_{FRB}}$ are each suited to a different class of KDE. $\Delta\mathrm{DM_{pulsar}}$ has smooth edges and can be adequately modelled with a Gaussian kernel and a fixed bandwidth. The sharp edge of $\Delta\mathrm{DM_{FRB}}$, however, necessitates a varying bandwidth and a kernel with a steep cut-off.

In § 3.1 we outline standard KDE and in § 3.2 we describe the modifications for asymmetric, bandwidth-varying KDE.

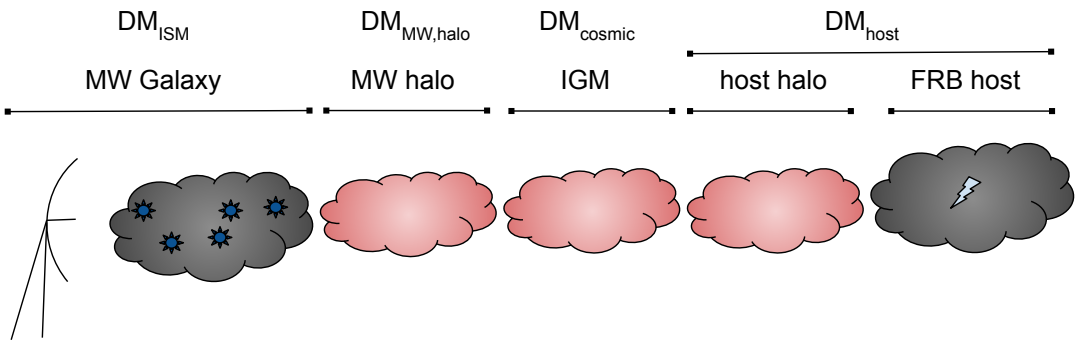

Figure 1. Schematic of the radio telescope (left-most image), the distribution of electrons (cloud shapes) that contribute to DM, and the millisecond transients (sun and lightning symbols) that are used to measure the DM. The regions shown in red have electrons, but no sources of millisecond transients. For sources distributed throughout their host galaxies and host galaxies distributed over a range of distances, the minimal Milky Way, IGM and FRB host DM contributions are zero.

3.1. *Standard KDEs*

Consider an independent and identically distributed sample $\{X_i : i = 1, ..., n\}$ drawn from some unknown distribution $f(x)$. We wish to obtain an estimate $\hat{f}(x)$ of this distribution using KDE:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(X_i - x\right) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \quad , \tag{3.1}$$

where K is the kernel and $h > 0$ is the bandwidth. The kernel is the underlying distribution function and the bandwidth is a smoothing parameter. In standard KDE symmetric kernels are used, such as Gaussian, triangular, cosine, biweight, triweight, or Epanechnikov. While an Epanechnikov kernel is most optimal in terms of the mean squared error, a Gaussian kernel is the most widely used: the loss of efficiency is marginal ($\sim 5\%$) and the distribution offers convenient mathematical properties. As such, a Gaussian kernel is used in our analysis of $\Delta\mathrm{DM}_{\mathrm{pulsar}}$. Bandwidth selection is a trade-off between the bias of the KDE and its variance. Often the bandwidth is chosen to minimize the mean integrated squared error (MISE),

$$\mathrm{MISE}(h) = \mathrm{E}\left[\int \left(\hat{f}(x) - f(x)\right)^2 dx\right] \quad , \tag{3.2}$$

which is equivalent to the expected L_2 risk function. $f(x)$ is unknown, however it can be approximated through various techniques (see Jones et al. (1996)). One can also use rule-of-thumb bandwidth estimators, such as Silverman's (Silverman 1986) and Scott's (Scott 1979), however these assume the underlying distribution is Gaussian. In our analysis we use `scikit-learn` to select the optimal bandwidth via cross-validation.

The `KernelDensity()` function invokes a nearest neighbors based approach: instead of using the full data set to estimate the density at each point, a number of neighboring points are selected based on the bandwidth. This improves the algorithm efficiency by ignoring distant points that have a negligible effect. KDEs are generated for a range of bandwidths, and `GridSearchCV()` is used to find the optimal bandwidth. Here n-fold cross-validation is performed. The pulsar data is divided into n subsets, a KDE is generated using the data from $n-1$ subsets (training data), and the performance of the KDE is evaluated on the remaining subset (test data) by calculating the log-likelihood, $\sum \log \hat{p}(x_i)$. This process is repeated n times, using a different subset as the test set each time, to give a final (averaged) log-likelihood score. In this manner, scores are calculated for a range of bandwidths. The bandwidth with the maximum log-likelihood is selected for the analysis ($h \approx 10$).

3.2. *Asymmetric KDEs*

Standard KDE performs well when the underlying distribution is unbounded and the density of data is relatively uniform. We will show, however, that the $\Delta\mathrm{DM}_{\mathrm{FRB}}$ distribution has data concentrated towards the front of the distribution and is bounded on $[0, \infty)$. This presents two problems that standard KDE cannot resolve. Firstly, a fixed bandwidth h entails a trade-off between large and small scale structure: over-dense regions will be over-smoothed by a large h, and under-dense regions will be over fitted if h is too small. Secondly, symmetric kernels have significant bias at or near a boundary, known as edge or boundary effects. A fixed and symmetric kernel will allocate weight outside of the density region when smoothing the distribution.

Various techniques have been developed that attempt to resolve this issue, eg. data reflection (Schuster 1985), boundary kernels (Müller 1991, 1993; Müller & Wang 1994), the hybrid method (Hall & Wehrly 1991), generating pseudo-data (Cowling & Hall 1996), data binning and local polynomial fitting (Cheng et al. 1997), and others. One can also invoke asymmetric kernels (such as gamma, lognormal and inverse Gaussian) and variable bandwidths. In this work we use gamma estimators developed by Chen (2000) and expanded upon by Jeon & Kim (2013) and Hoffmann & Jones (2015).

The gamma PDF with standard gamma function $\Gamma(\cdot)$ is given by

$$K_{k,\theta}(x) = \frac{x^{k-1}\exp(-\frac{x}{\theta})}{\theta^k \Gamma(k)} \quad , \tag{3.3}$$

with scale parameter k and shape parameter θ. Chen (2000) take $k = \rho_h(x)$ and $\theta = h$ with random gamma variables X_i to obtain

$$K_{\rho_h(x),h}(X_i) = \frac{X_i^{\rho_h(x)-1}\exp(-\frac{X_i}{h})}{h^{\rho_h(x)}\Gamma(\rho_h(x))} \quad , \tag{3.4}$$

with

$$\rho_h(x) = \begin{cases} \dfrac{x}{h}\,, & \text{if} \quad x \geq 2h \quad , \\ \left(\dfrac{x}{2h}\right)^2 + 1\,, & \text{if} \quad x \in [0, 2h) \quad . \end{cases}$$

The resulting gamma estimator is given by

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\rho_h(x),h}(X_i) \quad . \tag{3.5}$$

The shape of gamma kernels vary naturally, allowing for different smoothness at different points of the distribution. Further, because gamma kernels are non-negative, the gamma estimator itself is unlikely to deviate below zero. The bandwidths h depend either on the point of estimation ($h(x)$; a balloon estimator), or on the sample associated with a kernel ($h(X_i)$; sample-smoothing estimator). In this analysis we consider the former.

Another challenge for standard KDEs is that regions with few samples have overestimated densities and regions with many are underestimated. Shifted KDEs minimize this bias by moving samples from higher to lower density regions. Combining this with balloon estimators (Hoffmann & Jones 2015), one has

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\rho_h(x),h(x)}\left(X_i - h^p(x)\delta(x)\right) \quad , \tag{3.6}$$

where p is the order of the kernel and $\delta(x)$ is the shift. The kernel is shifted by $h^p(x)\delta(x)$, which vanishes for small bandwidths. For our analyses, we use Python code by Hoffmann & Jones (2015)[4], where the optimal bandwidth for each kernel is chosen by minimizing the MISE.

4. METHODOLOGY AND RESULTS

4.1. *Bounding the DM Distributions*

As described in § 2, we wish to estimate a maximum $\Delta\mathrm{DM}_{\mathrm{pulsar}}$ and a minimum $\Delta\mathrm{DM}_{\mathrm{FRB}}$ from the observed distributions. We will first apply the appropriate formalism to derive a PDF for each. The minimum/maximum of the PDF, however, is not a precisely posed quantity. Here we introduce a metric tailored primarily for $\Delta\mathrm{DM}_{\mathrm{FRB}}$ as an estimator after experimenting on simulated distributions (§ 4.3 and § A): the maximum gradient of the distributions, $\max\left[f'(\Delta\mathrm{DM})\right]$. This approach is based on the physical prior that the $\mathrm{DM}_{\mathrm{FRB}}$ distributions will have sharp cut-offs, which will hold if the variance in $\mathrm{DM}_{\mathrm{MW,halo}}$ is much less than its average. It is further supported by the current set of FRB observations. The observed $\Delta\mathrm{DM}_{\mathrm{pulsar}}$ PDF, on the other hand, is more evenly distributed with smoother edges. As such, estimates for $\max\left[\Delta\mathrm{DM}_{\mathrm{pulsar}}\right]$ given by the metric are more conservative. This effect is discussed in § A.

In § 4.2, KDE analysis is performed on observed transient samples to place current constraints on $\mathrm{DM}_{\mathrm{MW,halo}}$ from $\Delta\mathrm{DM}_{\mathrm{pulsar}}$ and $\Delta\mathrm{DM}_{\mathrm{FRB}}$. In 4.3, the KDE (gamma) methodology is analysed by simulating $\Delta\mathrm{DM}_{\mathrm{FRB}}$. Random samples of size $n = 100, 1000$ and 10,000 are taken and $\min\left[\Delta\mathrm{DM}_{\mathrm{FRB,sim}}\right]$ compared to the known inputs. This analysis also offers insight into the statistical power of future samples.

4.2. *Observed Sample*

To define our sample of pulsars and FRBs, we use the largest aggregation sites for each type of object. For pulsars, we downloaded the ATNF pulsar catalog (version 1.61; Manchester et al. 2005). For FRBs, we downloaded the FRBCat (downloaded 25 February 2020, verified events only; Petroff et al. 2016).

The Milky Way electron distribution is more complex at low Galactic latitudes owing to contributions from spiral arms, HII regions, and supernova remnants. Electron density models are most complex on size scales smaller than 200 pc and within 1 kpc of the Sun (Cordes & Lazio 2003). To minimize systematic error introduce by the model, we only consider sources more than

[4] Available at https://github.com/tillahoffmann/asymmetric_kde

200/1000 ≈ 20 deg from the galactic plane; we also compare the results with a second, more conservative cut to estimate systematic error. We also remove all pulsars within 5 deg of the Magellanic clouds. For a latitude limit of $|b| > 20\,\mathrm{deg}$, the samples include 371 pulsars and 83 FRBs. For a latitude limit of $|b| > 30\,\mathrm{deg}$, the samples include 215 pulsars and 64 FRBs. Owing to the significant decrease in FRB data for $|b| > 30\,\mathrm{deg}$, the final results presented in this paper use a Galactic cut of $|b| > 20\,\mathrm{deg}$.

This analysis requires correcting by the total $\mathrm{DM_{ISM}}$ contribution estimated from the Milky Way. Even at high Galactic latitudes, the electron density models have systematic uncertainties on the order of tens of percent due to modeling errors (Schnitzeler 2012). We estimate $\mathrm{DM_{ISM}}$ with both the NE2001 (Cordes & Lazio 2002, 2003) and YMW16 (Yao et al. 2017) models as a way of estimating potential systematic errors.

We then generated distributions of $\Delta\mathrm{DM_{pulsar}}$ and $\Delta\mathrm{DM_{FRB}}$, as given by Equations 2.3 and 2.6. These are shown in Figure 2a and 2b. As expected, the majority of $\Delta\mathrm{DM_{pulsar}}$ values are negative with a small tail to positive values. In contrast, the $\Delta\mathrm{DM_{FRB}}$ distribution is exclusively positive and rises sharply at $\Delta\mathrm{DM_{FRB}} \approx 64\,\mathrm{pc\,cm^{-3}}$.

We applied KDE (with Gaussian and gamma kernels, respectively) to the observed $\Delta\mathrm{DM_{pulsar}}$ and $\Delta\mathrm{DM_{FRB}}$ distributions to derive PDFs for each. The dark, thick curves in Figures 2a and 2b show the results. Also overlaid on the figures are a series of distributions derived from 1000 resampled data sets (100 shown). Table 2 reports the final results for both models on both Galactic latitude samples. In general, we find that the uncertainty on $\Delta\mathrm{DM_{FRB}}$ values are dominated by the size of the FRB sample. However, the uncertainty on the two distributions is largely insensitive to Galactic latitude cut. The YMW16 model tends to have slightly smaller $\mathrm{DM_{ISM}}$ values for this sample, which yields larger $\max[\Delta\mathrm{DM_{pulsar}}]$ and $\min[\Delta\mathrm{DM_{FRB}}]$ estimates. However, the separation of these distributions is not sensitive to the Galactic electron density model.

4.3. *Simulated Sample*

We now simulate $\Delta\mathrm{DM_{FRB}}$ to explore how the estimation of $\min[\Delta\mathrm{DM_{FRB}}]$ is likely to improve as more FRB data becomes available and to assess our choice of metric for $\min[\Delta\mathrm{DM_{FRB}}]$. From Equation 2.5,

$$\Delta\mathrm{DM_{FRB}} = \mathrm{DM_{MW,halo}} + \mathrm{DM_{cosmic}} + \mathrm{DM_{host}} \quad . \tag{4.1}$$

$\mathrm{DM_{MW,halo}}$ has a positive minimum, whereas $\mathrm{DM_{cosmic}}$ and $\mathrm{DM_{host}}$—in principle—have minimums of zero. As such, $\mathrm{DM_{MW,halo}}$ provides a zero-point offset for

	Latitude	$\max[\Delta\mathrm{DM_{pulsar}}]$	$\mathrm{DM_{MW,halo}}$
NE2001	$\lvert b\rvert > 20\,\mathrm{deg}$	$-2 \pm 2\,(\mathrm{stat}) \pm 9\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$	$> -11\,\mathrm{pc\,cm^{-3}}$
	$\lvert b\rvert > 30\,\mathrm{deg}$	$-4 \pm 3\,(\mathrm{stat}) \pm 8\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$	$> -13\,\mathrm{pc\,cm^{-3}}$
YMW16	$\lvert b\rvert > 20\,\mathrm{deg}$	$7 \pm 2\,(\mathrm{stat}) \pm 9\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$	$> -2\,\mathrm{pc\,cm^{-3}}$
	$\lvert b\rvert > 30\,\mathrm{deg}$	$4 \pm 2\,(\mathrm{stat}) \pm 8\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$	$> -5\,\mathrm{pc\,cm^{-3}}$

(a)

	Latitude	$\min[\Delta\mathrm{DM_{FRB}}]$	$\mathrm{DM_{MW,halo}}$
NE2001	$\lvert b\rvert > 20\,\mathrm{deg}$	$54^{+40}_{-19}\,(\mathrm{stat}) \pm 9\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$	$< 127\,\mathrm{pc\,cm^{-3}}$
	$\lvert b\rvert > 30\,\mathrm{deg}$	$45^{+39}_{-9}\,(\mathrm{stat}) \pm 7\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$	$< 110\,\mathrm{pc\,cm^{-3}}$
YMW16	$\lvert b\rvert > 20\,\mathrm{deg}$	$63^{+27}_{-21}\,(\mathrm{stat}) \pm 9\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$	$< 123\,\mathrm{pc\,cm^{-3}}$
	$\lvert b\rvert > 30\,\mathrm{deg}$	$52^{+37}_{-11}\,(\mathrm{stat}) \pm 7\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$	$< 113\,\mathrm{pc\,cm^{-3}}$

(b)

Table 2. Constraints derived from (a) pulsar and (b) FRB observations. NE2001 and YMW16 are used to model $\mathrm{DM_{ISM}}$ with $|b| > 20\,\mathrm{deg}$ and $|b| > 30\,\mathrm{deg}$. $\max[\Delta\mathrm{DM_{pulsar}}]$ and $\min[\Delta\mathrm{DM_{FRB}}]$ are calculated at 1σ, and upper and lower limits for $\mathrm{DM_{MW,halo}}$ at 95% c.l. . Systematic errors are taken to be the difference between NE2001 and YMW16 estimates. KDE with Gaussian kernels and fixed bandwidths are used to model $\Delta\mathrm{DM_{pulsar}}$, and KDE with gamma kernels and varying bandwidths are used to model $\Delta\mathrm{DM_{FRB}}$.

$\Delta\mathrm{DM_{FRB}}$, i.e. $\min[\Delta\mathrm{DM_{FRB}}] > 0$. For the following simulation, $\mathrm{DM_{MW,halo}}$ is chosen to be a delta function at $30\,\mathrm{pc\,cm^{-3}}$ and $\mathrm{DM_{host}}$ is approximated by a lognormal distribution with a mean of $\mu = 40\,\mathrm{pc\,cm^{-3}}$ and a standard deviation of $\sigma = 0.5$. Other models for these quantities are explored in § A.

To generate a cosmic DM contribution to the simulation, we must adopt a distribution of redshifts for the FRBs. We choose to estimate it from the observed $\mathrm{DM_{FRB}}$ values. Specifically, we adopt a DM–z relation[5] to convert the observed sample of $\mathrm{DM_{FRB}}$ values to a set of redshifts. Here the observed sample set has $|b| > 20\,\mathrm{deg}$ and $\mathrm{DM_{ISM}}$ is subtracted off with NE2001. We then applied standard KDE with a Gaussian kernel to build a PDF of the z values from which random draws may be taken. The draws are fed back into the DM–z relationship to obtain the average cosmic contribution to the DM,

$$\langle \mathrm{DM_{cosmic}}(z) \rangle = \int \frac{\bar{n}_e ds}{1+z} \quad , \tag{4.2}$$

where $\bar{n}_e = f_d(z)\rho_b(z)\mu_e/\mu_m m_p$ is the average electron density, f_d is the fraction of cosmic baryons in diffuse

[5] Code available at https://github.com/FRBs/FRB

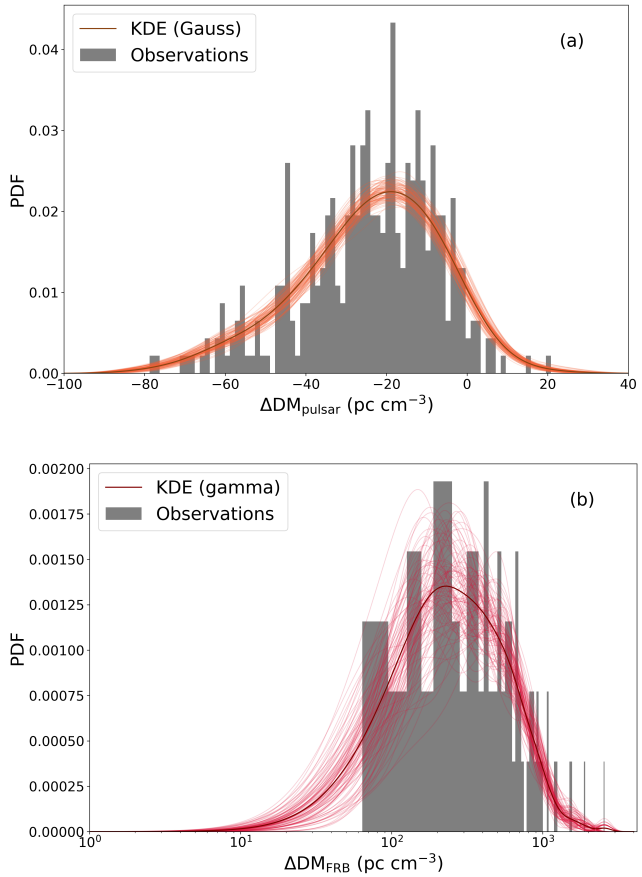

Figure 2. Distributions for observed samples, restricted to $|b| > 20\,\mathrm{deg}$ and using NE2001 for modeling $\mathrm{DM_{ISM}}$. Overlaid on the data are PDFs derived with KDE. (a) $\Delta\mathrm{DM_{pulsar}}$ KDEs (with Gaussian kernels and a fixed bandwidth) overlaid on the observed data. The dark orange curve denotes the PDF estimated with the original data, and the lighter curves denote PDFs generated with resampled data. The bandwidth for each distribution is selected with cross-correlation and a search range between $h = 8$ and $h = 15$. (b) $\Delta\mathrm{DM_{FRB}}$ KDEs (with gamma kernels and variable bandwidths) overlaid on the observed data. The thick dark red curve denotes the PDF generated with the original data and the lighter curves denote PDFs generated with the resampled data.

ionised gas, $\rho_b \equiv \Omega_b \rho_c$ is the cosmic baryonic mass density, and μ_m and μ_e describe properties of helium.

We allow for deviations of $\mathrm{DM_{cosmic}}$ from the average value following the formalism presented in Macquart & Ekers (2018). Our treatment is simpler than theirs; specifically, we assume that the fractional standard deviation of $\langle \mathrm{DM_{cosmic}} \rangle$ is $\sigma_{\mathrm{DM}} = F z^{-1/2}$ with $F = 0.2$. We may then generate a simulated $\mathrm{DM_{cosmic}}$ distribution based on the z distribution and random draws from a Gaussian characterized by $\sigma_{\mathrm{DM}} = 1$ and truncated at $\pm 1\sigma$. Throughout, we enforce $\mathrm{DM_{cosmic}} > 0$. The resultant $\mathrm{DM_{cosmic}}$ values are added to $\mathrm{DM_{halo}}$ and $\mathrm{DM_{host}}$ to give the simulated PDF of $\Delta\mathrm{DM_{FRB}}$.

Figure 3a shows a realization of this simulated PDF for $n = 10,000$ draws. This realization has an absolute minimum of $\Delta\mathrm{DM_{FRB}} = 30\,\mathrm{pc\,cm^{-3}}$ and rises sharply due to the host and $\mathrm{DM_{cosmic}}$ contributions. The dark red curve is the KDE (gamma) using the original data set and the other red curves are distributions generated with resampled data.

We explore the sensitivity of the analysis and results to samples size n as follows. For $n = 100, 1000$ and 10,000, we draw a random set of $\Delta\mathrm{DM_{FRB}}$ values and model the distributions with KDE (gamma). We then estimate a minimum value from the gradients of the PDFs, i.e. $\min[\mathrm{DM_{FRB}}]$ is the value which maximizes the slope of the KDE. Since each n PDF is complemented by 1000 PDFs resampled from the original data set, 1000 minima are available for error estimation. The distribution of $\min[\Delta\mathrm{DM_{FRB}}]$ values are shown in Table 3. As n increases, the dispersion in $\min[\Delta\mathrm{DM_{FRB}}]$ decreases and the central values approach $\approx 34\,\mathrm{pc\,cm^{-3}}$ (Figure 3b). Adding more than 10,000 samples has no notable effect on the results.

The simulation estimates are skewed to the left for small n and approach a Gaussian distribution with increased confidence as n increases (Figure 3b). While the mean values of the distributions are similar (Table 3), a sample size of $n = 100$ is inadequate to place a constraint with reasonable confidence. The confidence level does however improve significantly as n approaches 10,000.

Other choices for $\mathrm{DM_{host}}$ are explored to ensure the metric $\min[\Delta\mathrm{DM_{FRB}}] = \max[f'(\Delta\mathrm{DM_{FRB}})]$ is reasonably robust to changes in the FRB simulation. Results are consistent, as detailed in § A. The smoother the leading edge of $\Delta\mathrm{DM_{FRB}}$ i.e. the smoother $\mathrm{DM_{host}}$), the more conservative the limits become, and a very sharp edge for $\Delta\mathrm{DM_{FRB}}$ i.e. a delta function for $\mathrm{DM_{host}}$) is described well by the metric. These cases represent extreme examples of possible host galaxy DM distributions.

No. FRBs	$\min[\Delta\mathrm{DM_{FRB}}]$	$\mathrm{DM_{MW,halo}}$
100	$37 \pm 24\,(\mathrm{stat})\,\mathrm{pc\,cm^{-3}}$	$< 114\,\mathrm{pc\,cm^{-3}}$
1000	$35 \pm 7\,(\mathrm{stat})\,\mathrm{pc\,cm^{-3}}$	$< 55\,\mathrm{pc\,cm^{-3}}$
10000	$34 \pm 2\,(\mathrm{stat})\,\mathrm{pc\,cm^{-3}}$	$< 44\,\mathrm{pc\,cm^{-3}}$

Table 3. Simulation estimates for different sample sizes with $\min[\Delta\mathrm{DM_{FRB,sim}}] = 30\,\mathrm{pc\,cm^{-3}}$. The second column gives the recovered measurements for $\min[\Delta\mathrm{DM_{FRB}}]$ at 1σ and the last column gives an upper limit for $\mathrm{DM_{MW,halo}}$ (95% c.l.).

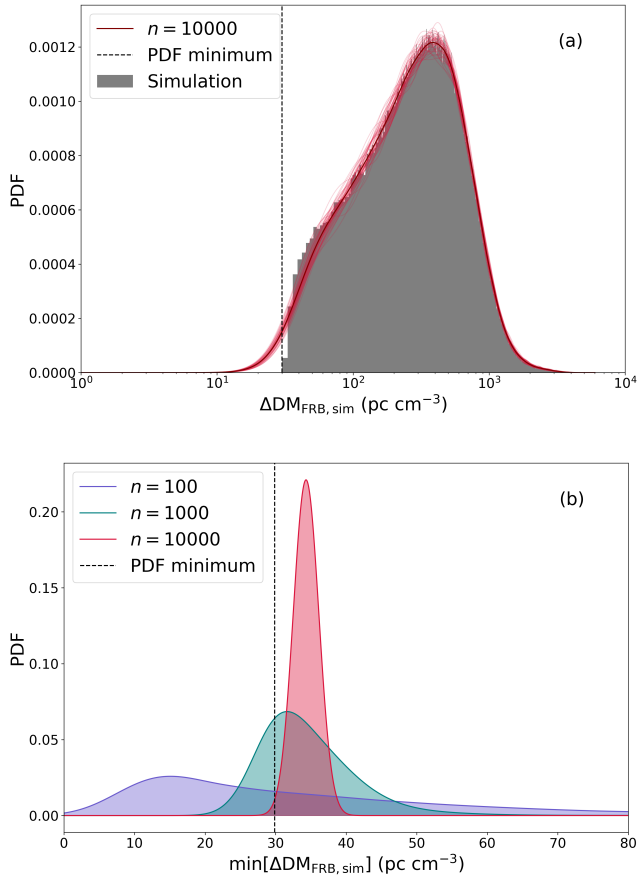

Figure 3. (a): Distribution of $\Delta\mathrm{DM}_{\mathrm{FRB,sim}}$ from simulated data. The KDE (gamma) estimation for $n = 10,000$ is denoted by the thicker dark red line. The thinner red lines show the ensemble of KDEs from resampled data. (b): Distributions of $\min[\Delta\mathrm{DM}_{\mathrm{FRB,sim}}]$ given by the maximum gradients of the KDE (gamma) PDFs. As the sample size increases, solutions settle with higher certainty to $\min[\Delta\mathrm{DM}_{\mathrm{FRB,sim}}] = 34\,\mathrm{pc\,cm^{-3}}$, which is $4\,\mathrm{pc\,cm^{-3}}$ above the absolute minimum.

5. DISCUSSION

The principle empirical result of our work is a conservative upper limit on the DM contribution of the Milky Way halo. At 1σ, $\mathrm{DM_{MW,halo}} = 63^{+27}_{-21}\,(\mathrm{stat}) \pm 9\,(\mathrm{sys})\,\mathrm{pc\,cm^{-3}}$ ($|b| > 20\,\mathrm{deg}$, YMW16). This can be converted to a conservative upper limit of $\mathrm{DM_{MW,halo}} < 123\mathrm{pc\,cm^{-3}}$ (95% c.l.). This includes the ISM and halo, and potentially a non-zero contribution from the FRB host galaxy, which is plausibly several tens $\mathrm{pc\,cm^{-3}}$ (see below). This limit also includes a non-zero contribution from the cosmic web ($\mathrm{DM_{cosmic}}$). That contribution is difficult to estimate at present but we note that the lowest redshift FRB ($z = 0.03$; Marcote et al. 2020) would yield an average $\mathrm{DM_{cosmic}}$ of $\approx 25\,\mathrm{pc\,cm^{-3}}$. A more realistic, yet speculative, upper limit to $\mathrm{DM_{MW,halo}}$ may therefore be $\approx 50\,\mathrm{pc\,cm^{-3}}$.

The results presented include two measurements of uncertainty: systematic uncertainties related to ISM models and statistical uncertainties related to the estimation techniques. Another point to consider is the effect that Galactic latitude has on results. Owing to the complexity of the electron distribution at lower Galactic latitudes, we consider cuts of $|b| > 20\,\mathrm{deg}$ and $|b| > 30\,\mathrm{deg}$. Results are largely insensitive to this cut, however the loss of data at $|b| > 30\,\mathrm{deg}$ (371 to 215 pulsars, and 83 to 64 FRBs), motivates a cut of $|b| > 20\,\mathrm{deg}$ for our final analysis.

Pulsar constraints are dominated by uncertainties in modeling $\mathrm{DM_{ISM}}$. We find that, on average, $\mathrm{DM_{ISM}}$ values recovered from NE2001 are $\approx 10\,\mathrm{pc\,cm^{-3}}$ lower than those from YMW16. Given the expectation that $\mathrm{DM_{MW,halo}} > 0$, we use YMW16 in our analysis (see Table 2b). This gives a final result of $\mathrm{DM_{MW,halo}} > -2\,\mathrm{pc\,cm^{-3}}$ (95% c.l.). We note that characterizing the line of sight to MW pulsars may help find HII regions that bias the $\mathrm{DM_{ISM}}$ estimate, allowing for improvement in the pulsar sample.

FRB constraints are predominantly limited by sample size n, i.e., our simulations show a significant improvement as n increases. For an absolute value of $\mathrm{DM_{MW,halo}} = 30\,\mathrm{pc\,cm^{-3}}$, limits for $n = 100$, 1000 and 10,000 are $\mathrm{DM_{MW,halo}} < 114\,\mathrm{pc\,cm^{-3}}$, $\mathrm{DM_{MW,halo}} < 55\,\mathrm{pc\,cm^{-3}}$, and $\mathrm{DM_{MW,halo}} < 44\,\mathrm{pc\,cm^{-3}}$, respectively (95% c.l.). This suggests that once thousands of FRBs have been observed, the constraints will greatly improve.

Even the conservative limit of $\mathrm{DM_{MW,halo}} < 123\,\mathrm{pc\,cm^{-3}}$ offers a valuable bound to models of the Galactic halo and the Local Group that our Galaxy resides within. Scenarios that adopt a Galactic halo mass $M_{\mathrm{halo}} \approx 10^{12.2}\,\mathrm{M_\odot}$ which has retained all of its cosmic average of baryons estimate $\mathrm{DM_{MW,halo}} > 50\,\mathrm{pc\,cm^{-3}}$ (Prochaska & Zheng (2019), but see Keating & Pen (2020)). Furthermore, models which would predict the gas traces the dark matter profile would yield $\mathrm{DM_{MW,halo}} > 200\,\mathrm{pc\,cm^{-3}}$ (Figure 4); these are ruled out by our FRB analysis, and also their over-estimated X-ray emission (e.g. Fang et al. 2015). Our results also place an upper bound on the average contribution from the Local Group medium, consistent with current estimates (Prochaska & Zheng 2019). Clearly, as the observed FRB sample increases—one expects a dramatic leap from the CHIME survey (CHIME/FRB Collaboration et al. 2018)—the resultant limits may well distinguish between models where the Galaxy has retained the majority of its baryons from those where they have been expelled.

To illustrate the potential constraints, Figure 4 shows a model-based estimate for $\mathrm{DM_{MW,halo}}$ for a dark mat-

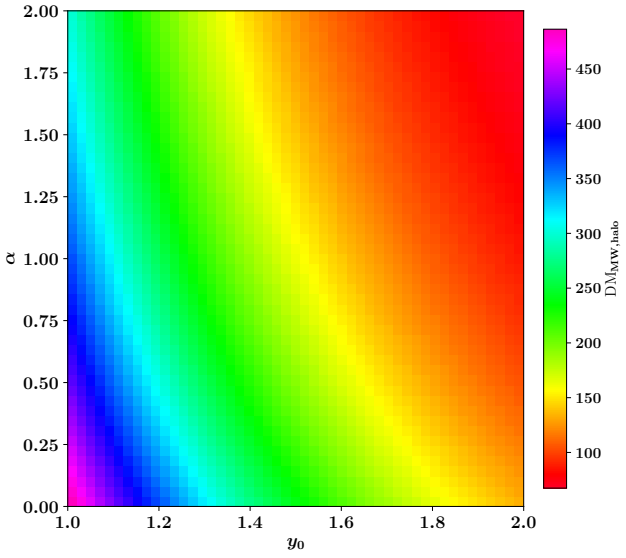

Figure 4. Predicted $DM_{MW,halo}$ for our Galaxy as a function of two shape parameters that describe the assumed baryonic density profile (Prochaska & Zheng 2019). The analysis assumes a Galactic halo with total baryonic mass $M_b \approx 2.4 \times 10^{11} M_\odot$ and that 75% of those baryons are in an ionized diffuse phase of the halo. The upper limit of $DM_{MW,halo} < 123\,pc\,cm^{-3}$ rules out density profiles that more closely resemble the NFW profile ($\alpha = 0, y_0 = 1$).

ter halo with mass $M_{halo} = 10^{12.2} M_\odot$, baryonic mass $M_b = \Omega_b/\Omega_m M_{halo} \approx 2.4 \times 10^{11} M_\odot$ and that 75% of those baryons are in a diffuse, ionized halo. The density profile is assumed to follow a modified Navarro-Frenk-White (NFW) profile parameterized by y_0 and α (see Mathews & Prochaska 2017; Prochaska & Zheng 2019). The upper limit to $DM_{MW,halo}$ estimated from our analysis prefers larger α, y_0 with a strict NFW profile ($\alpha = 0, y_0 = 1$) ruled out at high confidence unless $M_b \ll \Omega_b/\Omega_m M_{halo}$. Larger α, y_0 are inferred for our Galaxy and external ones from absorption-line analyses (e.g. Faerman et al. 2017; Mathews & Prochaska 2017).

We emphasize that ongoing FRB projects will offer complementary constraints on the magnitude and distribution of contributions from the host and the cosmic web to the upper limit on $DM_{MW,halo}$. In particular, well-localized FRBs reveal the host galaxy population and the redshift distribution of FRB events. From follow-up observations of the hosts, one may estimate the DM contribution from the host galaxy ISM through measurements of the Balmer line emission (e.g. Tendulkar et al. 2017; Chittidi et al. 2020). The two systems analyzed thus far yield $DM_{host,ISM} \approx 50$–$200\,pc\,cm^{-3}$. There are other FRBs (e.g. FRB 180924; Bannister et al. 2019) where the Balmer emission is low or even negligible at the FRB location and we infer $DM_{host,ISM} < 50\,pc\,cm^{-3}$. Within the next year, we expect to have a sample of ~ 20 hosts to derive the distribution.

One may additionally translate the estimated stellar mass of the host galaxy into a model-based estimate for the DM contribution from the halo gas of the host (Bannister et al. 2019; Prochaska & Zheng 2019). Current estimates range from $\approx 50\,pc\,cm^{-3}$ for the most massive hosts (Bannister et al. 2019) to $< 20\,pc\,cm^{-3}$ for FRB 181112 (Prochaska & Zheng 2019). From the redshift distribution of the localized FRBs, one may estimate the minimum typical contribution of DM_{cosmic} to the $DM_{MW,halo}$ limit. This bears an important caveat that the selection biases of the localized sample will not match those of the larger ensemble (e.g. due to differences in the radio frequencies and/or flux limit). One will need to account for these differences. Alternatively, one may focus on the analysis of the a localized sample alone once it grows to a sufficient sample size.

Last, we emphasize that other, future observations will also offer constraints on $DM_{MW,halo}$ independent of FRB analyses. We anticipate high-precision X-ray absorption-line spectroscopy of the Galactic halo from the upcoming Japanese XRISM mission. With a spectral resolution that will greatly exceed current X-ray satellites, the data will yield much more reliable estimates of O^{+5} and O^{+6} column densities across the sky. At the least, these yield conservative lower limits to $DM_{MW,halo}$. Another promising yet still unrealized opportunity is to discover pulsars in Andromeda or any other Local Group galaxy. These would offer a strict upper bound on $DM_{MW,halo}$ or even a well-informed value along that sightline.

6. CONCLUDING REMARKS

We have demonstrated how density estimation techniques can be used to probe the DM—i.e. the line-of-sight electron column density—of the MW Galactic halo. For the corrected DM_{pulsar} and DM_{FRB} distributions, we recover $\max[DM_{pulsar}] \approx 7 \pm 2\,(stat) \pm 9\,(sys)\,pc\,cm^{-3}$ and $\min[DM_{FRB}] \approx 63^{+27}_{-21}\,(stat) \pm 9\,(sys)\,pc\,cm^{-3}$ (1σ uncertainty). Conservative upper and lower limits on the Galactic halo dispersion measure are also derived: $DM_{MW,halo} > -2\,pc\,cm^{-3}$ and $DM_{MW,halo} < 123\,pc\,cm^{-3}$ (95% c.l.). Here the lower bound given by pulsars reflects only a fraction of the MW halo DM, and the upper bound given by FRBs includes a nominal contribution from the FRB host galaxy and IGM. In the latter case, the localization of FRBs at very low distances and/or on the outskirts of galaxies

would establish that the minimum DM would be more representative of the MW halo. Scenarios consistent with this include the collapse of compact objects (e.g. Falcke & Rezzolla 2013) that have been expelled from a host galaxy, as well as more exotic theories such as tiny electromagnetic explosions (which may occur in dark matter halos; Thompson 2017a,b) and cosmic strings (e.g. Vachaspati 2008; Yu et al. 2014; Zadorozhna 2015; Brandenberger et al. 2017).

We do not consider how $\mathrm{DM_{MW,halo}}$ may vary as a function of Galactic latitude. It may be possible with a sample of a couple thousand FRBs per region of sky, but is left to future work.

Our current estimates cannot yet discern whether the Milky Way has retained its cosmic average of baryons ($\mathrm{DM_{MW,halo}} > 50\,\mathrm{pc\,cm^{-3}}$), however in the near future, as more FRBs are reported, results may offer a valuable complement to other analyses. In the least, the methodology provides a reasonable—albeit conservative—estimate of $\mathrm{DM_{MW,halo}}$ and a minimum contribution from $\mathrm{DM_{host}}$. This may discern the viability of Galactic halo models and aid in the search for missing baryons.

ACKNOWLEDGMENTS

We would like to thank the anonymous referee for their insightful, thorough and valuable input. EP and JXP, as members of the Fast and Fortunate for FRB Follow-up team F^4, acknowledge support from NSF grant AST-1911140. CJL acknowledges support under NSF grant 2022546. This work was initiated as a project for the Kavli Summer Program in Astrophysics held at the University of California, Santa Cruz in 2019. The program was funded by the Kavli Foundation, The National Science Foundation, UC Santa Cruz, and the Simons Foundation. We thank them for their generous support. EP is supported by a L'Oréal-UNESCO For Women in Science Young Talents Fellowship, by a PhD fellowship from the South African National Institute for Theoretical Physics (NITheP), and by a top-up bursary from the South African Research Chairs Initiative of the Department of Science and Technology (SARChI) and the National Research Foundation (NRF) of South Africa. Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

REFERENCES

Bannister, K. W., Deller, A. T., Phillips, C., et al. 2019, Science, 365, 565, doi: 10.1126/science.aaw5903

Booth, C. M., Schaye, J., Delgado, J. D., & Dalla Vecchia, C. 2012, MNRAS, 420, 1053, doi: 10.1111/j.1365-2966.2011.20047.x

Boylan-Kolchin, M., Bullock, J. S., Sohn, S. T., Besla, G., & van der Marel, R. P. 2013, ApJ, 768, 140, doi: 10.1088/0004-637X/768/2/140

Brandenberger, R., Cyr, B., & Iyer, A. V. 2017. https://arxiv.org/abs/1707.02397

Bregman, J. N., Anderson, M. E., Miller, M. J., et al. 2018, ApJ, 862, 3, doi: 10.3847/1538-4357/aacafe

Caleb, M., Flynn, C., Bailes, M., et al. 2016, MNRAS, 458, 718, doi: 10.1093/mnras/stw109

Chen, S. X. 2000, Annals of the Institute of Statistical Mathematics, 52, 471, doi: 10.1023/A:1004165218295

Chen, W.-C., Tareen, A., & Kinney, J. B. 2018, Phys. Rev. Lett., 121, 160605, doi: 10.1103/PhysRevLett.121.160605

Cheng, M.-Y., Fan, J., & Marron, J. S. 1997, The Annals of Statistics, 25, 1691. http://www.jstor.org/stable/2959068

CHIME/FRB Collaboration, Amiri, M., Bandura, K., et al. 2018, ApJ, 863, 48, doi: 10.3847/1538-4357/aad188

Chittidi, J., Simha, S., Mannings, A., et al. 2020, ApJ, submitted

Coles, S. 2001, An Introduction to Statistical Modeling of Extreme Values (Springer Series in Statistics. Berlin: Springer-Verlag)

Cordes, J. M., & Chatterjee, S. 2019, ARA&A, 57, 417, doi: 10.1146/annurev-astro-091918-104501

Cordes, J. M., & Lazio, T. J. W. 2002. https://arxiv.org/abs/astro-ph/0207156

—. 2003. https://arxiv.org/abs/astro-ph/0301598

Cowling, A., & Hall, P. 1996, Journal of the Royal Statistical Society. Series B (Methodological), 58, 551. http://www.jstor.org/stable/2345893

Dai, X., Bregman, J. N., Kochanek, C. S., & Rasia, E. 2010, ApJ, 719, 119, doi: 10.1088/0004-637x/719/1/119

Faerman, Y., Sternberg, A., & McKee, C. F. 2013, ApJ, 777, 119, doi: 10.1088/0004-637X/777/2/119

—. 2017, ApJ, 835, 52, doi: 10.3847/1538-4357/835/1/52

Falcke, H., & Rezzolla, L. 2013, Astronomy and Astrophysics, 562, doi: 10.1051/0004-6361/201321996

Fang, T., Bullock, J. S., & Boylan-Kolchin, M. 2013, ApJ, 762, 20, doi: 10.1088/0004-637X/762/1/20

Fang, T., Buote, D. A., Bullock, J. S., & Ma, R. 2015, ApJS, 217, 21, doi: 10.1088/0067-0049/217/2/21

Faucher-Giguère, C.-A., & Kaspi, V. M. 2006, ApJ, 643, 332, doi: 10.1086/501516

Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, ApJ, 503, 518, doi: 10.1086/306025
Gaensler, B. M., Madsen, G. J., Chatterjee, S., & Mao, S. A. 2008, PASA, 25, 184, doi: 10.1071/AS08004
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 1995, Chapman & Hall, London
Hall, P., & Wehrly, T. E. 1991, Journal of the American Statistical Association, 86, 665, doi: 10.1080/01621459.1991.10475092
Henley, D. B., Shelton, R. L., Kwak, K., Joung, M. R., & Mac Low, M.-M. 2010, ApJ, 723, 935, doi: 10.1088/0004-637X/723/1/935
Hoffmann, T., & Jones, N. S. 2015. https://arxiv.org/abs/1512.03188
Jeon, Y., & Kim, J. H. 2013, Insurance: Mathematics and Economics, 53, 569, doi: 10.1016/j.insmatheco.2013
Jones, M. C., Marron, J. S., & Sheather, S. J. 1996, Journal of the American Statistical Association, 91, 401. http://www.jstor.org/stable/2291420
Keating, L. C., & Pen, U.-L. 2020, MNRAS, submitted
Kinney, J. B. 2014, Phys. Rev. E, 90, 011301, doi: 10.1103/PhysRevE.90.011301
—. 2015, Phys. Rev. E, 92, 032107, doi: 10.1103/PhysRevE.92.032107
Kocz, J., Ravi, V., Catha, M., et al. 2019, MNRAS, 489, 919, doi: 10.1093/mnras/stz2219
Kovács, O. E., Bogdán, Á., Smith, R. K., Kraft, R. P., & Forman, W. R. 2019, ApJ, 872, 83, doi: 10.3847/1538-4357/aaef78
Law, C. J., Bower, G. C., Burke-Spolaor, S., et al. 2018, ApJS, 236, 8, doi: 10.3847/1538-4365/aab77b
Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, Science, 318, 777, doi: 10.1126/science.1147532
Macquart, J.-P. 2018, Nature Astronomy, 2, 836, doi: 10.1038/s41550-018-0625-7
Macquart, J.-P., & Ekers, R. 2018, MNRAS, 480, 4211, doi: 10.1093/mnras/sty2083
Manchester, R. N., Fan, G., Lyne, A. G., Kaspi, V. M., & Crawford, F. 2006, ApJ, 649, 235, doi: 10.1086/505461
Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993, doi: 10.1086/428488
Marcote, B., Nimmo, K., Hessels, J. W. T., et al. 2020, Nature, 577, 190, doi: 10.1038/s41586-019-1866-z
Mathews, W. G., & Prochaska, J. X. 2017, The Astrophysical Journal, 846, L24, doi: 10.3847/2041-8213/aa8861
Müller, & Wang, J.-L. 1994, Biometrics, 50, 61. http://www.jstor.org/stable/2533197
Müller, H. 1991, Biometrika, 78, 521, doi: 10.1093/biomet/78.3.521
—. 1993, Scandinavian Journal of Statistics, 20, 313. http://www.jstor.org/stable/4616287
Petroff, E., Hessels, J. W. T., & Lorimer, D. R. 2019, Astron. Astrophys. Rev., 27, 4, doi: 10.1007/s00159-019-0116-6
Petroff, E., Barr, E. D., Jameson, A., et al. 2016, PASA, 33, e045, doi: 10.1017/pasa.2016.35
Prochaska, J. X., Weiner, B., Chen, H.-W., Mulchaey, J., & Cooksey, K. 2011, ApJ, 740, 91, doi: 10.1088/0004-637X/740/2/91
Prochaska, J. X., & Zheng, Y. 2019, MNRAS, 485, 648, doi: 10.1093/mnras/stz261
Prochaska, J. X., Macquart, J.-P., McQuinn, M., et al. 2019, Science, doi: 10.1126/science.aay0073
Ravi, V., et al. 2019, Nature, 572, 352, doi: 10.1038/s41586-019-1389-7
Ridley, J. P., Crawford, F., Lorimer, D. R., et al. 2013, MNRAS, 433, 138, doi: 10.1093/mnras/stt709
Riihimaki, J., & Vehtari, A. 2014, Bayesian Anal., 9, 425, doi: 10.1214/14-BA872
Salem, M., Besla, G., Bryan, G., et al. 2015, ApJ, 815, 77, doi: 10.1088/0004-637X/815/1/77
Schnitzeler, D. H. F. M. 2012, MNRAS, 427, 664, doi: 10.1111/j.1365-2966.2012.21869.x
Schuster, E. 1985, Communications in Statistics - Theory and Methods, 14, 1123, doi: 10.1080/03610928508828965
Scott, D. W. 1979, Biometrika, 66, 605, doi: 10.1093/biomet/66.3.605
Silverman, B. W. 1986, Density Estimation for Statistics and Data Analysis (London: Chapman & Hall)
Skare, Ø., Bølviken, E., & Holden, L. 2003, Scandinavian Journal of Statistics, 30, 719, doi: 10.1111/1467-9469.00360
Tendulkar, S. P., et al. 2017, ApJ., 834, L7, doi: 10.3847/2041-8213/834/2/L7
Thompson, C. 2017a, ApJ, 844, 65, doi: 10.3847/1538-4357/aa7684
—. 2017b, ApJ, 844, 162, doi: 10.3847/1538-4357/aa7845
Vachaspati, T. 2008, Phys. Rev. Lett., 101, 141301, doi: 10.1103/PhysRevLett.101.141301
White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341, doi: 10.1093/mnras/183.3.341
Yamasaki, S., & Totani, T. 2020, The Astrophysical Journal, 888, 105, doi: 10.3847/1538-4357/ab58c4
Yao, J. M., Manchester, R. N., & Wang, N. 2017, ApJ, 835, 29, doi: 10.3847/1538-4357/835/1/29
Yu, Y.-W., Cheng, K.-S., Shiu, G., & Tye, H. 2014, JCAP, 1411, 040, doi: 10.1088/1475-7516/2014/11/040

Zadorozhna, L. V. 2015, Advances in Astronomy and Space Physics, 5, 43, doi: 10.17721/2227-1481.5.43-50

APPENDIX

A. MINIMUM OF FRB DM DISTRIBUTION

We postulate that the minimum of the $\Delta\mathrm{DM_{FRB}}$ distribution can be approximated by $\min[\Delta\mathrm{DM_{FRB}}] = \max[f'(\Delta\mathrm{DM_{FRB}})]$. This metric is based on the prior that the underlying distribution has a sharp leading edge and is motivated by simulations. To bear weight, the metric must hold for a wide range of reasonable $\Delta\mathrm{DM_{FRB}}$ distributions.

The MW can be given by a delta function (as its DM is thought to vary by 10 $\mathrm{pc\,cm^{-3}}$) and the cosmic DM distribution can be modelled theoretically. The distribution of host galaxy DMs, however, is unknown. In the main analysis we consider a lognormal distribution with $\mu = 40\ \mathrm{pc\,cm^{-3}}$ and a standard deviation of $\sigma = 0.5$. Here we consider two extreme variations: a delta function at $30\mathrm{pc\,cm^{-3}}$ and a broad Gaussian distribution with $\mu = 60\,\mathrm{pc\,cm^{-3}}$ and $\sigma = 0.5$. The former distribution makes the edge of $\Delta\mathrm{DM_{FRB}}$ sharper and the latter makes it smoother. The metric is a reasonable approximation for the combined $\mathrm{DM_{MW,halo}} + \mathrm{DM_{host}}$ contribution when each distribution is sharp (Figure 5a). When $\mathrm{DM_{host}}$ has a smooth edge, the estimates are more conservative (Figure 5b). Thus, provided the leading edge of $\Delta\mathrm{DM_{FRB}}$ is sufficiently sharp, the metric for determining the distribution minimum can be considered reasonably robust.

Looking at Figure 5, a sample size of $n = 1000$ appears sufficient to provide an estimate consistent with that of $n = 10,000$. For $n = 100$, distributions are wide and skewed to the left, providing results that are clearly premature.

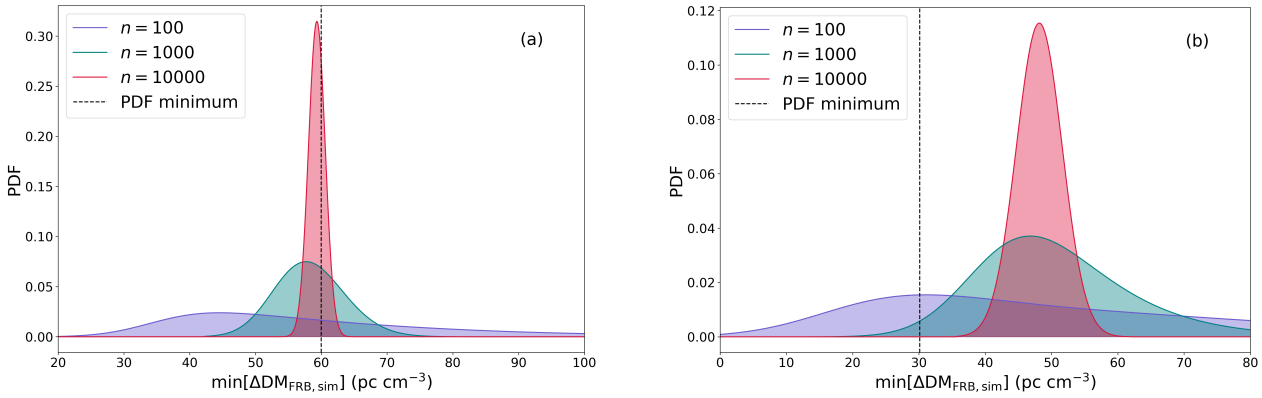

Figure 5. (a) $\min[\Delta\mathrm{DM_{FRB}}]$ with $\mathrm{DM_{host}}$ a delta function at 30 $\mathrm{pc\,cm^{-3}}$. The absolute minimum is 60 $\mathrm{pc\,cm^{-3}}$. (b) $\min[\Delta\mathrm{DM_{FRB}}]$ for a Gaussian $\mathrm{DM_{host}}$ with $\mu = 60\,\mathrm{pc\,cm^{-3}}$ and $\sigma = 15$. The absolute minimum is 30 $\mathrm{pc\,cm^{-3}}$.

B. DENSITY ESTIMATION USING FIELD THEORY

Density estimation using field theory (DEFT; Kinney 2014, 2015; Chen et al. 2018) is a newly developed technique specifically developed for the small data regime. It takes a Bayesian field theory approach to density estimation in small data sets using a Laplace approximation of the Bayesian posterior (also see Riihimaki & Vehtari (2014)). An advantage of DEFT over standard density estimation methods is that the method does not require the manual identification of critical parameters nor does it require the specification of boundary conditions. The DEFT simulations in this paper use the Python package `SUFTware` (Statistics Using Field Theory) by Chen et al. (2018).

Consider n data points $(x_1, x_2, ..., x_n)$ drawn from a known probability distribution $Q_{\mathrm{true}}(x)$ with x intervals of length L. We wish to find the best estimate $Q^*(x)$ of this distribution and the accompanying ensemble of other plausible estimates. Each distribution $Q(x)$ is parameterized by a real field $\phi(x)$, ensuring that $Q(x)$ is positive and normalized:

$$Q(x) = \frac{e^{-\phi(x)}}{\int dx' e^{-\phi(x')}} \quad . \tag{B1}$$

Using scalar field theory, a prior $p(\phi|\ell)$ is formulated that favours smooth probability densities. Specifically, Kinney (2015) consider priors of the form

$$p(\phi|\ell) = \frac{e^{-S_\ell^0[\phi]}}{Z_\ell^0} \quad , \tag{B2}$$

with action

$$S_\ell^0[\phi] = \int \frac{dx}{L} \frac{\ell^{2\alpha}}{2} \left(\partial^\alpha \phi\right)^2 \quad , \tag{B3}$$

and partition function

$$Z_\ell^0 = \int \mathcal{D}\phi e^{-S_\ell^0[\phi]} \quad . \tag{B4}$$

Here, ℓ gives the length scale below which ϕ fluctuations are strongly damped and $\alpha > 0$ is an integer in the range [1,...,4] that determines the smoothness. The resultant posterior is given by

$$p(\phi|\text{data}, \ell) = \frac{e^{-S_\ell[\phi]}}{Z_\ell} \quad , \tag{B5}$$

with nonlinear action

$$S_\ell[\phi] = \int \frac{dx}{L} \left\{ \frac{\ell^2 \alpha}{2} \left(\partial^\alpha \phi\right)^2 + nLR\phi + ne^{-\phi} \right\} \quad , \tag{B6}$$

and partition function

$$Z_\ell = \int \mathcal{D}\phi e^{-S_\ell[\phi]} \quad . \tag{B7}$$

$R(x) = \frac{1}{n}\sum_{i=1}^{n} \partial(x - x_i)$ is a histogram that summarizes the data.

Maximum *a posteriori* (MAP) density estimation approximates the posterior $p(\phi|\text{data}, \ell)$ as a δ function given by the mode of the posterior, at which the action $S_\ell[\phi]$ is then minimized. It has been shown that even without imposing boundary conditions on ϕ, $S_\ell[\phi]$ has a unique minimum (Kinney 2015). The optimal length scale ℓ^* is identified by maximizing the Bayesian evidence $p(\text{data}|\ell)$.

The uncertainty in the DEFT estimate Q^* is determined by sampling the Bayesian posterior,

$$p(Q|\text{data}) = \int dl p(\ell|\text{data}) p(Q|\text{data}, \ell) \quad , \tag{B8}$$

by first drawing ℓ from $p(\ell|\text{data})$ and then drawing Q from $p(Q|\text{data}, \ell)$. Laplace approximation is used to estimate $p(Q|\text{data}, \ell)$ by constructing a Gaussian centered at its MAP value. This gives the Laplace posterior,

$$p_{\text{Lap}}(Q|\text{data}) = \int dl p(\ell|\text{data}) p_{\text{Lap}}(Q|\text{data}, \ell) \quad , \tag{B9}$$

from which an ensemble of distributions Q can be sampled. Some of the Qs, however, are clearly not representative of the underlying distribution. Importance resampling is thus used to remove unfavorable distributions, where each ϕ is given a weight,

$$w_\ell[\phi] = \exp\left(S_\ell^{\text{Lap}}[\phi] - S_\ell[\phi]\right) \quad , \tag{B10}$$

proportional to its probability of being drawn (Chen et al. 2018). DEFT uses importance resampling with replacement, however for this work we invoke importance resampling without replacement.

When a posterior turns out to be a poor approximation of the target distribution, a few of the sampled distributions are given very large weights and the majority are given small weights (Gelman et al. 1995; Skare et al. 2003). When resampling with replacement, the heavily weighted distributions become significantly over represented. In our case, $\sim$ 60–70% of the sampled distributions were duplicates, which lead to notable bias when calculating the upper and lower bounds of $\text{DM}_{\text{MW,halo}}$. As such, we use a set of the most probable distributions, with limited replications. Specifically, we select 500 out of 1000 distributions via importance sampling without replacement. This lowered the duplication rate to $\sim$ 10%.

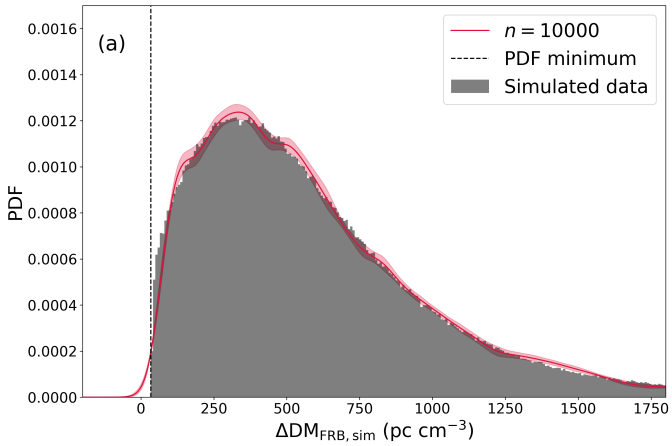

Figure 6. Distributions of ΔDM_{FRB} for 10,000 samples, restricted to $|b| > 20$ deg and using NE2001 for modeling DM_{ISM}. Overlaid on the data are PDFs derived with DEFT. The thick line denotes the DEFT Bayesiean posterior and shaded line denotes standard deviation of the set of PDFs derived by sampling the Bayesiean posterior.

We approximate the FRB distribution described in §4.3 using DEFT for $n = 100$, $n = 1000$ and $n = 10,000$. Even for large n DEFT is unable to adequately describe the sharp edge of the simulated distribution. In Figure 6a, the PDF tail extends below zero, violating the physical condition that $\Delta DM_{FRB} > 0$. Further, the PDF cuts straight through the front of the simulated distribution and so bypasses the structure of the distribution's edge.

C. GENERALIZED EXTREME VALUE

A standard statistical technique for estimating the maximum values of an ensemble to fit it with a Generalized Extreme Value (GEV) PDF (e.g. Coles 2001). This technique, however, is most applicable for assessing the upper limit of a distribution with a long tail. For ΔDM_{FRB}, this holds for the largest values but the lowest values rise sharply as one may expect from the MW and host contributions.

Nevertheless, we attempted to estimate the minimum of ΔDM_{FRB} following the standard practice of assessing the maximum of the negative of the distribution (Coles 2001). The results reported a minimum value at effectively infinite confidence at the lowest ΔDM_{FRB} in the distribution and we found the results were unstable to random sampling.